Quantiles, expectiles and splines

DeRossi, Giuliano; Harvey, Andrew

Empfohlene Zitierung / Suggested Citation:
DeRossi, G., & Harvey, A. (2009). Quantiles, expectiles and splines. *Journal of Econometrics*, *152*(2), 179-185. https://doi.org/10.1016/j.jeconom.2009.01.001

Accepted Manuscript

Quantiles, expectiles and splines

Giuliano De Rossi, Andrew Harvey

PII: S0304-4076(09)00015-3
DOI: 10.1016/j.jeconom.2009.01.001
Reference: ECONOM 3149

To appear in: *Journal of Econometrics*

Please cite this article as: De Rossi, G., Harvey, A., Quantiles, expectiles and splines. *Journal of Econometrics* (2009), doi:10.1016/j.jeconom.2009.01.001

Quantiles, Expectiles and Splines

Giuliano De Rossi and Andrew Harvey*

Faculty of Economics, Cambridge University

August 20, 2008

Abstract

A time-varying quantile can be fitted by formulating a time series model for the corresponding population quantile and iteratively applying a suitably modified state space signal extraction algorithm. It is shown that such quantiles satisfy the defining property of fixed quantiles in having the appropriate number of observations above and below. Like quantiles, time-varying expectiles can be estimated by a state space signal extraction algorithm and they satisfy properties that generalize the moment conditions associated with fixed expectiles. Because the state space form can handle irregularly spaced observations, the proposed algorithms can be adapted to provide a viable means of computing spline-based non-parametric quantile and expectile regressions.

KEYWORDS: Asymmetric least squares; cubic splines; quantile regression; signal extraction; state space smoother.

JEL Classification: C14, C22

*Corresponding author: Faculty of Economics and Politics, Sidgwick Avenue, Cambridge CB3 9DD. Telephone +44-(0) 1223 335228.

Email: ach34@econ.cam.ac.uk.

1 Introduction

The movements in a time series may be described by time-varying quantiles. These may be estimated non-parametrically by fitting a simple moving average or a more elaborate kernel. An alternative approach is to formulate a partial model, the role of which is to focus attention on some particular feature - here a quantile - so as to provide a (usually nonlinear) weighting of the observations that will extract that feature by taking account of the dynamic properties of the series. The model is not intended to be taken as a full description of the distribution of the observations. Indeed models for different features, for example different quantiles, may not be consistent with each other.

In an earlier paper, we showed how time-varying quantiles could be fitted to a sequence of observations by setting up a state space model and iteratively applying a suitably modified signal extraction algorithm; see De Rossi and Harvey (2006). Here we determine the conditions under which a linear time series model for the quantile will satisfy the defining property of fixed quantiles in having the appropriate number of observations above and below.

Expectiles are similar to quantiles except that they are defined by tail expectations; see Newey and Powell (1987). Here we show how time-varying expectiles can be estimated by a state space signal extraction algorithm. This is similar to the algorithm used for quantiles, but estimation is more straightforward and much quicker. We then show that the conditions needed for a time-varying expectile to generalize the moment conditions associated with fixed expectiles are similar to those needed for a time-varying quantile to satisfy the defining property of fixed quantiles.

Section 2 reviews the ideas underlying fixed quantiles and expectiles. Section 3 then describes the signal extraction algorithms for estimating them when they are time-varying and establishes some basic properties. The final part of the paper is concerned with non-parametric estimation of regression models using splines. It has long been known that cubic splines can be fitted by signal extraction procedures because the state space form can handle irregularly spaced observations from a continuous time model. The proposed algorithms for time-varying quantiles and expectiles are easily adapted so as to provide a viable means of computing spline-based non-parametric quantile and expectile regressions. As well as illustrating the technique, we give a general proof of the equivalence between splines and the continuous time models underlying our signal extraction procedures for quantiles and expectiles.

2 Quantiles and expectiles

Let $\xi(\tau)$ - or, when there is no risk of confusion, ξ - denote the $\tau - th$ quantile. The probability that an observation is less than $\xi(\tau)$ is τ, where $0 < \tau < 1$.

Given a set of T observations, $y_t, t = 1, .., T$, (which may be from a cross-section or a time series), the sample quantile, $\widetilde{\xi}(\tau)$, can be obtained by sorting the observations in ascending order. However, it is also given as the solution to minimizing

$$S_\tau = \sum_{t=1}^{T} \rho_\tau(y_t - \xi) = \sum_{y_t < \xi} (\tau - 1)(y_t - \xi) + \sum_{y_t \geq \xi} \tau(y_t - \xi) \tag{1}$$

with respect to ξ, where $\rho_\tau(.)$ is the *check function,* defined for quantiles as

$$\rho_\tau(y_t - \xi) = (\tau - I(y_t - \xi < 0))\,(y_t - \xi) \tag{2}$$

and $I(.)$ is the indicator function.

Expectiles, denoted $\mu(\omega), 0 < \omega < 1$, are similar to quantiles but they are determined by tail expectations rather than tail probabilities. For a given value of ω, the sample expectile, $\widetilde{\mu}(\omega)$, is obtained by minimizing the asymmetric least squares function,

$$S_\omega = \sum \rho_\omega(y_t - \mu) = \sum |\omega - I(y_t - \mu < 0)|\,(y_t - \mu)^2, \tag{3}$$

with respect to μ. Differentiating S_ω and dividing by minus two gives

$$\sum_{t=1}^{T} |\omega - I(y_t - \mu < 0)|\;\;(y_t - \mu). \tag{4}$$

The sample expectile, $\widetilde{\mu}(\omega)$, is the value of μ that makes (4) equal to zero. Setting $\omega = 0.5$ gives the mean, that is $\widetilde{\mu}(0.5) = \overline{y}$. For other ω's it is necessary to iterate.

3 Signal extraction

A framework for estimating time-varying quantiles, $\xi_t(\tau)$, can be set up by assuming that they are generated by stochastic processes and are connected to the observations through a measurement equation

$$y_t = \xi_t(\tau) + \varepsilon_t(\tau), \qquad t = 1, ..., T, \tag{5}$$

where $\Pr(\varepsilon_t(\tau) < 0) = \tau$ with $0 < \tau < 1$. The disturbances, $\varepsilon_t(\tau)$, are assumed to be serially independent and independent of $\xi_t(\tau)$. The problem is then one of signal extraction. The assumption that the quantile or expectile follows a stochastic process can be regarded as a device for inducing local weighting of the observations. One possibility is a random walk,

$$\xi_t(\tau) = \xi_{t-1}(\tau) + \eta_t(\tau), \qquad \eta_t(\tau) \backsim IID(0, \sigma^2_{\eta(\tau)}). \tag{6}$$

A smoother quantile can be extracted by a local linear trend

$$\begin{aligned} \xi_t &= \xi_{t-1} + \beta_{t-1} + \eta_t \\ \beta_t &= \beta_{t-1} + \zeta_t \end{aligned} \tag{7}$$

where β_t is the slope and ζ_t is $IID(0, \sigma^2_\zeta)$. It is well known that in a Gaussian model setting $Var(\eta_t) = \sigma^2_\zeta/3$ and $Cov(\eta_t, \zeta_t) = \sigma^2_\zeta/2$ results in the smoothed estimates being a cubic spline.

The model for expectiles is set up in a similar way with (5) replaced by $y_t = \mu_t(\omega) + \varepsilon_t(\omega)$ where the ω-expectile of $\varepsilon_t(\omega)$ is equal to zero.

3.1 Theory and computation

The state space form (SSF) for a univariate time series is:

$$\begin{aligned} y_t &= \mathbf{z}_t'\boldsymbol{\alpha}_t + \varepsilon_t, \qquad Var(\varepsilon_t) = \sigma_t^2, \qquad t = 1, ..., T \\ \boldsymbol{\alpha}_t &= \mathbf{T}_t\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t, \qquad Var(\boldsymbol{\eta}_t) = \mathbf{Q}_t \end{aligned} \tag{8}$$

where $\boldsymbol{\alpha}_t$ is an $m \times 1$ state vector, $\mathbf{z}_t$ is a non-stochastic $m \times 1$ vector, σ_t^2 is a non-negative scalar, $\mathbf{T}_t$ is an $m \times m$ non-stochastic transition matrix and $\mathbf{Q}_t$ is an $m \times m$ covariance matrix. The specification is completed by assuming that $\boldsymbol{\alpha}_1$ has mean $\mathbf{a}_{1|0}$ and covariance matrix $\mathbf{P}_{1|0}$ and that the serially independent disturbances ε_t and $\boldsymbol{\eta}_t$ are independent of each other and of the initial state.

Consider the criterion function

$$\begin{aligned} J &= -\sum_{t=1}^{T} h_t^{-1}\rho(y_t - \mathbf{z}_t'\boldsymbol{\alpha}_t) - \frac{1}{2}\sum_{t=2}^{T}(\boldsymbol{\alpha}_t - \mathbf{T}_t\boldsymbol{\alpha}_{t-1})'\,\mathbf{Q}_t^{-1}\,(\boldsymbol{\alpha}_t - \mathbf{T}_t\boldsymbol{\alpha}_{t-1}) \\ &- \frac{1}{2}(\boldsymbol{\alpha}_1 - \mathbf{a}_{1|0})'\mathbf{P}_{1|0}^{-1}(\boldsymbol{\alpha}_1 - \mathbf{a}_{1|0}), \end{aligned} \tag{9}$$

where $\rho(y_t - \mathbf{z}_t'\boldsymbol{\alpha}_t)$ is as in (2) or (3), with $\mathbf{z}_t'\boldsymbol{\alpha}_t$ equal to $\xi_t(\tau)$ or $\mu_t(\omega)$, $\mathbf{Q}_t$ and $\mathbf{P}_{1|0}$ are are assumed positive definite matrices as in (8) and h_t is a non-stochastic sequence of positive scalars. For example, in the local linear trend case (7) $\boldsymbol{\alpha}_t = (\xi_t, \beta_t)'$ and $\mathbf{z}' = (1\ 0)$, while $\mathbf{T}$ is upper triangular with nonzero elements equal to one. Suppose that the initial state and the $\boldsymbol{\eta}_t's$ are normally distributed. For a Gaussian model of the form (8) the logarithm of the joint density of the observations and the states is, ignoring irrelevant terms, given by J with $\rho(y_t - \mathbf{z}_t'\boldsymbol{\alpha}_t) = (y_t - \mu_t(0.5))^2$ and $h_t = 2\sigma_t^2$. Differentiating J with respect to to each element of $\boldsymbol{\alpha}_t$ gives a set of equations, which, when set to zero and solved, gives the minimum mean square error estimates of $\boldsymbol{\alpha}_t$. These they may be computed efficiently by the Kalman filter and associated smoother (KFS) as described in Durbin and Koopman (2001, pp. 70-73). If all the elements in the state are nonstationary and given a diffuse prior, the last term in J disappears. An algorithm is available as a subroutine in the SsfPack set of programs within Ox; see Koopman et al. (1999).

We can think of (9) as a criterion function that provides the basis for computing a quantile or expectile subject to a set of constraints imposed by the time series model for the quantile or expectile[1]. For expectiles differentiating J gives

$$\begin{aligned}
\frac{\partial J}{\partial \boldsymbol{\alpha}_1} &= \mathbf{z}_1(2/h_1)IE(y_1 - \mathbf{z}_1'\boldsymbol{\alpha}_1) - \mathbf{P}_{1|0}^{-1}(\boldsymbol{\alpha}_1 - \mathbf{a}_{1|0}) + \mathbf{T}_2'\mathbf{Q}_2^{-1}(\boldsymbol{\alpha}_2 - \mathbf{T}_2\boldsymbol{\alpha}_1) \\
\frac{\partial J}{\partial \boldsymbol{\alpha}_t} &= \mathbf{z}_t(2/h_t)IE(y_t - \mathbf{z}_t'\boldsymbol{\alpha}_t) - \mathbf{Q}_t^{-1}(\boldsymbol{\alpha}_t - \mathbf{T}_t\boldsymbol{\alpha}_{t-1}) + \mathbf{T}_{t+1}'\mathbf{Q}_{t+1}^{-1}(\boldsymbol{\alpha}_{t+1} - \mathbf{T}_{t+1}\boldsymbol{\alpha}_t), \\
& \quad t=2,\dots,T-1, \\
\frac{\partial J}{\partial \boldsymbol{\alpha}_T} &= \mathbf{z}_T(2/h_T)IE(y_T - \mathbf{z}_T'\boldsymbol{\alpha}_T) - \mathbf{Q}_T^{-1}(\boldsymbol{\alpha}_T - \mathbf{T}_T\boldsymbol{\alpha}_{T-1}) \qquad (10)
\end{aligned}$$

where

$$IE(y_t - \mu_t(\omega)) = |\omega - I(y_t - \mu_t(\omega) < 0)| \; (y_t - \mu_t(\omega)), \qquad t = 1, ..., T. \quad (11)$$

The smoothed estimates, $\widetilde{\boldsymbol{\alpha}}_t$, satisfy the equations obtained by setting these derivatives equal to zero. Let $h_t = g_t/\kappa$, where κ is a constant, the interpretation of which will become apparent in sub-section 3.3. For any expectile,

[1] It could also be regarded as the log of the joint density of a model where the measurement error is an asymmetric double exponential (quantile) or asymmetric normal (expectile). But such a model could not be taken seriously as a data generating process.

adding and subtracting $\mathbf{z}_t g_t^{-1} \mathbf{z}_t' \boldsymbol{\alpha}_t$ to the equations in (10) allows the first term to be written as

$$\mathbf{z}_t g_t^{-1}[\mathbf{z}_t'\boldsymbol{\alpha}_t + 2\kappa IE\left(y_t - \mathbf{z}_t'\boldsymbol{\alpha}_t\right)] - \mathbf{z}_t g_t^{-1}\mathbf{z}_t'\boldsymbol{\alpha}_t, \qquad t = 1, \ldots, T. \tag{12}$$

This suggests that we set up an iterative procedure in which the estimate of the state at the i-th iteration, $\widehat{\boldsymbol{\alpha}}_t^{(i)}$, is computed from the KFS applied to a set of synthetic 'observations' constructed as

$$\widehat{y}_t^{(i-1)} = \mathbf{z}_t'\widehat{\boldsymbol{\alpha}}_t^{(i-1)} + 2\kappa IE\left(y_t - \mathbf{z}_t'\widehat{\boldsymbol{\alpha}}_t^{(i-1)}\right). \tag{13}$$

The iterations are carried out until the $\widehat{\boldsymbol{\alpha}}_t^{(i)\prime}s$ converge whereupon $\widetilde{\mu}_t(\omega) = \mathbf{z}_t'\widetilde{\boldsymbol{\alpha}}_t$.

For quantiles, the first term in each of the three equations of (10) is given by $\mathbf{z}_t h_t^{-1} IQ(y_t - \mathbf{z}_t'\boldsymbol{\alpha}_t)$, where

$$IQ(y_t - \xi_t(\tau)) = \begin{cases} \tau - 1, & \text{if} \quad y_t < \xi_t(\tau) \\ \tau, & \text{if} \quad y_t > \xi_t(\tau) \end{cases} \qquad t = 1, ..., T, \tag{14}$$

and the synthetic observations in the KFS are

$$\widehat{y}_t^{(j-1)} = \mathbf{z}_t'\widehat{\boldsymbol{\alpha}}_t^{(j-1)} + \kappa IQ\left(y_t - \mathbf{z}_t'\widehat{\boldsymbol{\alpha}}_t^{(j-1)}\right), \qquad t = 1, ..., T \tag{15}$$

However, the possibility of a solution where the estimated quantile passes through an observation means that the algorithm has to be modified somewhat; see De Rossi and Harvey (2006).

3.2 Properties

Estimates of time-varying quantiles and expectiles obtained from the smoothing equations of the previous sub-section can be shown to satisfy properties that generalize the defining characteristics of fixed quantiles and expectiles. In order to establish the conditions under which these properties hold, we first prove a preliminary result for any time series model in SSF, (8). It is assumed that the state has been arranged so that the first element represents the level of the quantile or expectile and that (without loss of generality) the first element in $\mathbf{z}_t$ has been set to unity. Let the sum of the first derivatives, with respect to $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_T$, of the second term of J be written $\mathbf{j}_2' = \sum_{t=1}^{T} \mathbf{A}_t \boldsymbol{\alpha}_t$, where the $\mathbf{A}_t's$ are $m \times m$ matrices.

Lemma *For a model in SSF with a diffuse prior on the initial state, a sufficient condition for the first element in the vector $\mathbf{j}_2'$ to be zero is that the first column of $T_t - I$ consists of zeroes for all $t = 2, ..., T$.*

Proof - Summing the terms in the derivatives in question gives

$$\begin{aligned}
\sum \mathbf{A}_t \boldsymbol{\alpha}_t &= (\mathbf{Q}_2^{-1}\mathbf{T}_2 - \mathbf{T}_2'\mathbf{Q}_2^{-1}\mathbf{T}_2)\boldsymbol{\alpha}_1 \\
&+ \sum_{t=2}^{T-1} (\mathbf{Q}_{t+1}^{-1}\mathbf{T}_{t+1} - \mathbf{T}_{t+1}'\mathbf{Q}_{t+1}^{-1}\mathbf{T}_{t+1} + \mathbf{T}_t'\mathbf{Q}_t^{-1} - \mathbf{Q}_t^{-1})\boldsymbol{\alpha}_t \\
&+ \left(\mathbf{T}_T'\mathbf{Q}_T^{-1} - \mathbf{Q}_T^{-1}\right)\boldsymbol{\alpha}_T \qquad (16)
\end{aligned}$$

The matrix associated with $\boldsymbol{\alpha}_1$ is $\mathbf{A}_1 = \mathbf{Q}_2^{-1}\mathbf{T}_2 - \mathbf{T}_2'\mathbf{Q}_2^{-1}\mathbf{T}_2 = (\mathbf{I} - \mathbf{T}_2')\mathbf{Q}_2^{-1}\mathbf{T}_2$. A sufficient condition for it to have a null first row is that $\mathbf{I} - \mathbf{T}_2'$ has a null first row. The matrix associated with $\boldsymbol{\alpha}_T$ is $(\mathbf{T}_T' - \mathbf{I})\mathbf{Q}_T^{-1}$ and the condition for it to have a null first row is that $\mathbf{T}_T' - \mathbf{I}$ has a null first row. On examining the matrices, $\mathbf{A}_t, t = 2, .., T-1$, associated with the remaining state vectors we see that an analogous condition is sufficient for each to have a null first row. Letting some of the states have proper priors does not affect the result as long as they are uncorrelated with the diffuse prior on the first element in the state.

Proposition 1 *If the condition of the Lemma holds, then, for expectiles, the generalized moment condition*

$$\sum_{t=1}^{T} |\omega - I(y_t - \mathbf{z}_t' \widetilde{\boldsymbol{\alpha}}_t)| \, (y_t - \mathbf{z}_t' \widetilde{\boldsymbol{\alpha}}_t)/h_t = 0,$$

holds.

The result follows because, when the first element in the vector $\sum \mathbf{A}_t \boldsymbol{\alpha}_t$ is zero, differentiating the first term in J gives

$$\sum_{t=1}^{T} h_t^{-1} IE(y_t - \mathbf{z}_t' \widetilde{\boldsymbol{\alpha}}_t) = 0.$$

To give some intuition, in the special case of time invariant h_t Proposition 1 implies that the *weighted* sum of residuals is zero. When h_t is time invariant and $\omega = 0.5$ the solution is the Kalman filter and smoother and the sum of residuals is equal to zero.

The results for quantiles require a little more work.

Proposition 2 *If h_t is time-invariant and the conditions of the Lemma hold, the estimated quantiles satisfy the fundamental property of sample time-varying quantiles, namely that the number of observations that are less than the corresponding quantile, that is $y_t < \widetilde{\xi}_t(\tau)$, is no more than $[T\tau]$ while the number greater is no more than $[T(1-\tau)]$.*

Proof - Suppose that the only one point at which the quantile passes through an observation is at $t = s$, so $\widetilde{\xi}_s = y_s$. All the derivatives of J, defined in (9) with $\rho_\tau(.)$ as in (2), can be set to zero apart from this one. However, a small increase in $\widetilde{\xi}_s$ gives $IQ(y_s - \widetilde{\xi}_s)$ a value of $\tau - 1$ while a small decrease makes it equal to τ. Thus to have

$$IQ(y_s - \widetilde{\xi}_s) + \sum_{t \neq s} IQ(y_t - \widetilde{\xi}_t) = 0$$

implies

$$-\tau \leq \sum_{t \neq s} IQ(y_t - \widetilde{\xi}_t) \leq 1 - \tau.$$

When the quantile passes through k observations, a similar argument leads to

$$-k\tau \leq \sum_{t \notin C} IQ(y_t - \widetilde{\xi}_t) \leq k\,(1 - \tau) \tag{17}$$

where C is the set of all points such that $\widetilde{\xi}_s = y_s$. Now suppose that $\underline{n}$ denotes the number of observations (strictly) below the corresponding quantile while $\overline{n} = (T - \underline{n} - k)$ is the number (strictly) above. Then, abbreviating $IQ(y_t - \widetilde{\xi}_t)$ to IQ_t,

$$\sum_{t \notin C} IQ_t = \underline{n}\,(\tau - 1) + (T - \underline{n} - k)\,\tau = T\tau - \underline{n} - k\tau$$

Now $\sum_{t \notin C} IQ_t \geq -k\tau$ implies $\underline{n} \leq [\tau T]$ because $\sum_{t \notin C} IQ_t$ would be less than $-k\tau$ if $\underline{n}$ were greater than $[\tau T]$. Similarly, $\sum_{t \notin C} IQ_t \leq k(1 - \tau)$ implies $\overline{n} \leq [(1 - \tau)T]$ because $\sum_{t \notin C} IQ_t = \overline{n} - (1 - \tau)\,T + k\,(1 - \tau)$ would be greater than $k(1 - \tau)$ if $\overline{n}$ were to exceed $[(1 - \tau)T]$.

Proposition 3 *If h_t is not time-invariant, the estimated quantiles satisfy a* *generalization of the fundamental property, which is that*

$$\sum_{t\in B} 1/h_t \leq \tau \sum_{t=1}^{T} 1/h_t \qquad \textit{and} \qquad \sum_{t\in A} 1/h_t \leq (1-\tau)\sum_{t=1}^{T} 1/h_t$$

where $t \in B$ denotes the set of observations below the corresponding quantile and $t \in A$ denotes the set above.

The result follows because corresponding to (17) we have

$$-\tau \sum_{t\in C} 1/h_t \leq \sum_{t\notin C} (1/h_t) IQ(y_t - \mathbf{z}_t'\widetilde{\boldsymbol{\alpha}}_t) \leq (1-\tau)\sum_{t\in C} 1/h_t \tag{18}$$

The condition of the lemma is obviously satisfied by the random walk. It is also satisfied by the local linear trend. In a model with fixed explanatory variables, $\mathbf{x}_t$, the first equation in (8) becomes

$$y_t = \mathbf{x}_t'\boldsymbol{\delta} + \mathbf{z}_t'\boldsymbol{\alpha}_t + \varepsilon_t, \qquad t = 1, ..., T \tag{19}$$

and if the coefficient vector is put in the state vector as $\boldsymbol{\delta}_t$ and given a diffuse prior, the conditions apply to the transition equation for $\boldsymbol{\alpha}_t$ as before.

Finally we turn to the conditions under which quantiles and expectiles match up when they are time-varying.

Proposition 4 *If the distribution of y is time invariant when adjusted for changes in location and scale, and is continuous with finite mean*[2]*, the population $\tau-$quantiles and $\omega-$expectiles coincide for ω satisfying*

$$\omega = \frac{\int_{-\infty}^{\xi(\tau)} (y-\xi(\tau))dF(y)}{\int_{-\infty}^{\xi(\tau)} (y-\xi(\tau))dF(y) - \int_{\xi(\tau)}^{\infty} (y-\xi(\tau))dF(y)}$$

[2]Newey and Powell (1987, theorem 1) show that the expectiles are uniquely defined if the mean, $E(y) = \mu(0.5)$, exists.

where $F(y)$ is the cdf of y. Assuming this to be the case, $\widetilde{\mu}_t(\omega)$ is an estimator of the $\widetilde{\tau}-$quantile, $\xi_t(\widetilde{\tau})$, where $\widetilde{\tau}$ is defined as the proportion of observations for which $y_t < \widetilde{\mu}_t(\omega), t = 1, ..., T$.

When used in this way we will denote the estimator $\widetilde{\mu}_t(\omega)$ as $\widetilde{\mu}_t(\widetilde{\tau})$. However, it will not, in general, coincide with the time-varying $\widetilde{\tau}-$quantile estimated directly since it weights the observations differently. In particular, it is unlikely to pass through any observations.

3.3 Parameter estimation

The smoothing algorithms of sub-section 3.1 depend on parameters that can be estimated by cross validation. For time-varying quantiles, the function to be minimized is

$$CV(\tau) = \sum_{t=1}^{T} \rho_\tau(y_t - \widetilde{\xi}_t^{(-t)}) \tag{20}$$

where $\widetilde{\xi}_t^{(-t)}$ is the smoothed value at time t when y_t is dropped; see De Rossi and Harvey (2006). A similar criterion, $CV(\omega)$, may be used for expectiles.

In a time invariant model with quantiles or expectiles following a random walk, $\mathbf{Q}_t$ is a scalar equal to $\sigma^2_{\eta(\tau)}$ or $\sigma^2_{\eta(\omega)}$. We would like a suitable parameterization in terms of a quasi signal-noise ratio that is scale invariant. For the mean it will be recalled that $h_t = 2\sigma_t^2$ and so in a time invariant model the usual signal-noise ratio,[3] σ^2_η/σ^2, implies that $g_t = \sigma^2$ and $\kappa = 0.5$ in (12). A similar normalization can be applied for other expectiles so the

[3]In terms of the notation used in (8), $\sigma_t^2 = \sigma^2$ and $Q_t = \sigma_\eta^2\ \forall t$.

quasi signal-noise ratios are $q_\omega = \sigma^2_{\eta(\omega)}/\sigma^2$. Hence the iterations are based on (12) with g_t set to one and $\kappa = 0.5$.

For quantiles, σ^2_η/h is not scale invariant. We therefore consider the quasi signal-noise ratio, $q_\tau = \sigma^2_\eta/g$, with g defined so that q_τ is scale invariant. Since the variance is not robust it is better to estimate the inter-quartile range, r, and set g equal to its square. If the median is time-varying, it is estimated by setting $g_t = \kappa = 1$ and r is estimated from the residuals; the estimated quasi signal-noise may then be divided by the square of the estimate of r so as to make it scale invariant. For the other quantiles the iterative scheme is applied with g set to one and $\kappa = \widehat{r}$.

4 Nonparametric regression with cubic splines

A slowly changing quantile can be estimated by minimizing the criterion function $\sum \rho_\tau\{y_t - \xi_t\}$ subject to smoothness constraints. The cubic spline solution seeks to do this by finding a solution to

$$\min \sum_{t=1}^{T} \rho_\tau\{y_t - \xi(x_t)\} + \lambda_2 \left(\int \{\xi''(x)\}^2 dx \right) \tag{21}$$

where $\xi(x)$ is a continuous function with square integrable second derivative, $0 \le x \le T$ and $x_t = t$. The parameter λ_2 controls the smoothness of the spline. We show in the appendix that the same cubic spline is obtained by quantile signal extraction of (7) with $\lambda_2 = h/2\sigma^2_\zeta$. A random walk corresponds to $\xi'(x)$ rather than $\xi''(x)$ in the above formula; compare Kohn, Ansley and Wong (1992). Our proof not only shows that the well-known connection between splines and stochastic trends in Gaussian models carries over to quantiles, but it does so in a way that yields a more compact proof for the Gaussian case and shows that the result holds for expectiles. We furthermore establish the existence and uniqueness of the solution.

The SSF allows irregularly spaced observations to be handled since it can deal with systems that are not time invariant. The form of such systems is the implied discrete time formulation of a continuous time model; see Harvey (1989, p 487). This generalisation allows the handling of nonparametric quantile and expectile regression by cubic splines when there is only one explanatory variable. The observations, which may be from a cross-section, are arranged so that the values of the explanatory variable are in ascending order. Other variables can be included if they enter linearly as in (19).

Bosch et al. (1995) propose a solution to cubic spline quantile regression that uses quadratic programming. Unfortunately this necessitates the repeated inversion of large matrices of dimension up to $4T \times 4T$. This is very time consuming. Our signal extraction appears to be much faster (and more general) and makes estimation of the smoothing parameter (quasi signal-noise ratio) a feasible proposition.

The fundamental property of quantiles continues to hold with irregularly spaced observations. All that happens is that the SSF becomes time-varying. If there are multiple observations at some points then n, the total number of observations, replaces T, number of distinct points, in the summation. The proof follows by adding more $\rho(.)$ terms at times where there are multiple observations.

Proposition 5 *If n denotes the total number of observations while T is the number of distinct points at which observations occur, the fundamental property of quantiles is stated in terms of n rather than T.*

The only difference in the proof is in the summation involving $IQ(.)'s$ which now becomes

$$\sum_{j=1}^{n} h_{t(j)}^{-1} IQ(y_j - \mathbf{z}_{t(j)}' \widetilde{\boldsymbol{\alpha}}_{t(j)})$$

where $t(j)$ denotes that the $j-th$ observation is observed at point $t = t(j)$, $t \in \{1, \ldots, T\}$.

Proposition 6 *If there are multiple observations at some points, the generalized moment condition for expectiles is*

$$\sum_{j=1}^{n} h_{t(j)}^{-1} \left| \tau - I(y_j - \mathbf{z}_{t(j)}' \widetilde{\boldsymbol{\alpha}}_{t(j)}) \right| (y_j - \mathbf{z}_{t(j)}' \widetilde{\boldsymbol{\alpha}}_{t(j)}) = 0$$

An example of cubic spline regression is provided by the "motorcycle data", which records measurements of the acceleration, in milliseconds, of the head of a dummy in motorcycle crash tests. The data set was originally analysed by Silverman (1985) and has been used in a number of textbooks, including Koenker (2005, p 222-6). The observations are irregularly spaced and at some time points there are multiple observations. Harvey and Koopman (2000) highlight the stochastic trend connection.

Figure 1 shows the cubic spline expectiles obtained using the value of $\sigma_\zeta^2/\sigma^2 = 0.07$ computed by CV for the mean. (The ML estimate is 0.03). Although the expectiles lack the nice interpretation of quantiles, the graph

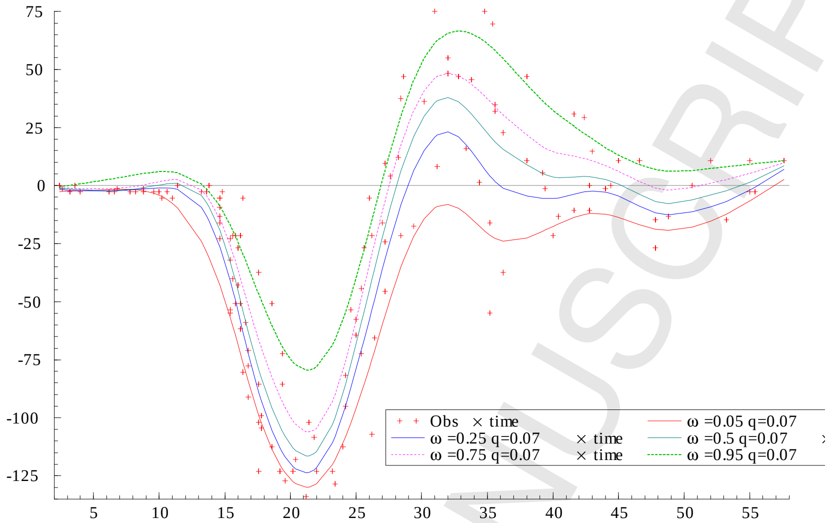

Figure 1: Cubic spline expectiles fitted to the motorcycle data. The parameter q_μ is estimated by cross validation.

gives a clear visual impression of the movements in level and dispersion. Of course if we count the number of observations below each expectile, they can be interpreted as quantiles if we are prepared to assume that the shape of the distribution is time invariant.

5 Conclusions

Time-varying quantiles and expectiles are of interest in themselves and provide information on various aspects of a time series, such as dispersion and asymmetry.

Time-varying quantiles can be fitted iteratively applying a suitably modified state space signal extraction algorithm. The algorithm for time-varying expectiles is much faster as there is no need to take account of corner solutions. We derive the conditions under which time-varying quantiles satisfy the defining property of fixed quantiles in having the appropriate number of observations above and below it, while expectiles satisfy properties that generalize the moment conditions associated with fixed expectiles.

Our model-based approach means that time-varying quantiles and expectiles can be used for forecasting. As such they offer an alternative to methods such as those in Engle and Manganelli (2004) and Granger and Sin (2000), that are based on conditional autoregressive models.

Finally we prove that if the underlying time series model is a Wiener process or an integrated Wiener process, then the solution for quantiles and expectiles is equivalent to fitting a spline; for an integrated Wiener process this is a cubic spline. We furthermore establish the existence and uniqueness of the solution. Because the state space form can handle irregularly spaced observations, the proposed algorithms are easily adapted to provide a viable means of computing spline-based non-parametric quantile and expectile regressions. We demonstrated how this worked for the 'motorcycle' data and showed, in that case, that fitting cubic spline expectiles gave a clear visual impression of the changing distribution.

APPENDIX

State space representation of quantile and expectile regression with smoothing splines

Consider a set of n observations $(y_1, \ldots, y_n)$ obtained at times $(t_1, \ldots, t_n)$, where $0 \leq t_1 < \ldots < t_n \leq b$. Moreover, consider the loss function ρ_τ defined in (2). We will deal with the problem of finding the function $f : [0, b] \to \mathbb{R}$ that minimises

$$\lambda_m \int_0^b \left[f^{(m)}(t)\right]^2 \mathrm{d}t + \sum_{i=1}^n \rho_\tau\left(y_i - f(t_i)\right) \tag{22}$$

for given $\tau \in (0, 1)$ and m, over all functions f having $m - 1$ absolutely continuous derivatives and square integrable $m-$th derivative.

Now consider the time series representation obtained by assuming that:

- $\left[f(0), f'(0), \ldots, f^{(m-1)}(0)\right] \sim \mathcal{N}(\mathbf{0}, \kappa \mathbf{I}_m)$;
-
$$f(t) = \sum_{j=1}^m \frac{t^{j-1}}{(j-1)!} f^{(j-1)}(0) + \sigma_w \int_0^t \frac{(t-s)^{m-1}}{(m-1)!} \mathrm{d}W_s \tag{23}$$
  where W_t is a Wiener process (in terms of the notation of sub-section 3.1, $\sigma_w = \sigma_\eta$ for $m = 1$ and σ_ζ for $m = 2$);
- the distribution of $y_i | f(t_i)$ is asymmetric double exponential, i.e. its pdf is
$$p\left(y_i | f(t_i)\right) \propto \exp\left[-\lambda^{-1} \rho_\tau\left(y_i - f(t_i)\right)\right], \tag{24}$$
  where λ is a constant.

Define $\mathbf{y} = (y_1, \ldots, y_n)'$ and $\mathbf{f} = (f(t_1), \ldots, f(t_n))$. We will show that, if $\lambda_m = \lambda/(2\sigma_w^2)$, as $\kappa \to \infty$ the mode of the smoothing distribution $p(\mathbf{f}|\mathbf{y})$ converges to the point $(f(t_1), \ldots, f(t_n))$ obtained by evaluating the solution of the problem (22) at $(t_1, \ldots, t_n)$.

Remark 7 *The quantile regression with smoothing splines problem described by Bosch et al. (1995) is a special case with* $m = 2$*. The result can easily be extended to expectile regression by replacing* $\rho_\tau(x)$ *with* $\rho_\omega(x) = |\omega - I(x < 0)| x^2$ *in the proof. This results in an asymmetric Gaussian distribution for the observations conditional on the signal.*

Remark 8 *If the density in the measurement equation (24) were Gaussian our argument would provide an alternative proof of the result of Wahba (1978) for the special cases* $m = 1, 2$*. This follows on noting that in a Gaussian model conditional means and conditional modes coincide. Wahba's proof requires the explicit solution of the spline smoothing problem (derived in Kimeldorf and Wahba, 1971), which is shown to be equal to the conditional mean. Our proof simply shows that the two optimisation problems, i.e. finding the mode and finding the optimal spline, are equivalent.*

Remark 9 *The existence and uniqueness of the solution to problem (22) depend crucially on the convexity of* $\rho_\tau(x)$*. This follows immediately from the fact that if* ρ_τ *is convex then the log-likelihood of the time series representation is strictly concave in* $\mathbb{R}^n$*.*

Proof

The mode of $p(\mathbf{f}|\mathbf{y})$ is found by solving $\max_{\mathbf{f}} p(\mathbf{f}|\mathbf{y})$.This is equivalent to solving $\max_{\mathbf{f}} p(\mathbf{y}, \mathbf{f})$ and we proceed by first noting that

$$p(\mathbf{y}, \mathbf{f}) = p(\mathbf{y}|\mathbf{f})\, p(\mathbf{f}). \tag{25}$$

Consider the joint distribution $p(\mathbf{f})$. It is a multivariate normal distribution with mean zero (because $f(0), f'(0), \ldots$ have zero mean) and covariance matrix $\sigma^2 \mathbf{W}_n + \kappa \mathbf{T}\mathbf{T}'$, where

$$\mathbf{T}' = \begin{bmatrix} 1 & \ldots & 1 \\ t_1 & \ldots & t_n \\ \vdots & & \vdots \\ t_1^{m-1}/(m-1)! & \ldots & t_n^{m-1}/(m-1)! \end{bmatrix}$$

and

$$\mathbf{W}_n = \mathrm{Cov}\left[\left(\int_0^{t_1} \frac{(t_1-s)^{m-1}}{(m-1)!} dW_s, \ldots, \int_0^{t_n} \frac{(t_n-s)^{m-1}}{(m-1)!} dW_s\right)'\right].$$

It can be easily shown that, for an $(n \times m)$ matrix a and a nonsingular $(n \times n)$ matrix Ω:

$$(I_n + aa')^{-1} = I_n - a(I_m + a'a)^{-1} a'$$

and

$$\Omega^{-1} - \Omega^{-1} a \left(a'\Omega^{-1}a\right)^{-1} a'\Omega^{-1} = a_\perp \left(a_\perp' \Omega a_\perp\right)^{-1} a_\perp',$$

where $a_\perp$ is an $n \times (n-m)$ matrix whose columns are orthogonal to the columns of a.

The above identities imply that

$$\begin{aligned} \lim_{\kappa\to\infty} \left(\sigma_w^2 \mathbf{W}_n + \kappa \mathbf{T}\mathbf{T}'\right)^{-1} &= \lim_{\kappa\to\infty} \sigma_w^{-2} \left(\mathbf{W}_n^{-1} - \mathbf{W}_n^{-1}\mathbf{T}\left(\mathbf{I}_{n-m}\frac{\sigma_w^2}{\kappa} + \mathbf{T}'\mathbf{W}_n^{-1}\mathbf{T}\right)^{-1} \mathbf{T}'\mathbf{W}_n^{-1}\right) \\ &= \sigma_w^{-2}\left(\mathbf{W}_n^{-1} - \mathbf{W}_n^{-1}\mathbf{T}\left(\mathbf{T}'\mathbf{W}_n^{-1}\mathbf{T}\right)\mathbf{T}'\mathbf{W}_n^{-1}\right) \\ &= \sigma_w^{-2}\mathbf{U}\left(\mathbf{U}'\mathbf{W}_n\mathbf{U}\right)^{-1}\mathbf{U}' \end{aligned}$$

where $\mathbf{U}'$ is a $(n-m)\times n$ matrix whose rows are orthogonal to the rows of $\mathbf{T}'$. As a result, the density of $\mathbf{f}$ becomes

$$\lim_{\kappa\to\infty} p(\mathbf{f}) \propto \exp\left(-\frac{1}{2\sigma_w^2}\mathbf{f}'\mathbf{U}\left(\mathbf{U}'\mathbf{W}_n\mathbf{U}\right)^{-1}\mathbf{U}'\mathbf{f}\right).$$

From (25) and (24) we have

$$\lim_{\kappa\to\infty} p(\mathbf{f}|\mathbf{y}) \propto \exp\left(-\frac{1}{2\sigma_w^2}\mathbf{f}'\mathbf{U}\left(\mathbf{U}'\mathbf{W}_n\mathbf{U}\right)^{-1}\mathbf{U}'\mathbf{f} - \frac{1}{\lambda}\sum_{i=1}^{n}\rho_\tau\left(y_i - f(t_i)\right)\right).$$

For $m=1$ we can set

$$\begin{aligned}
\mathbf{U}' &= \begin{bmatrix} \mathbf{u}_1' \\ \mathbf{u}_2' \\ \vdots \\ \mathbf{u}_{n-1}' \end{bmatrix}, \quad \mathbf{u}_i' = \left(\underbrace{0,\ldots,0}_{\text{i-1 zeros}}, \alpha_0^i, \alpha_1^i, \underbrace{0,\ldots,0}_{\text{n-1-i zeros}}\right), \\
\alpha_0^i &= \frac{1}{t_{i+1}-t_i} \\
\alpha_1^i &= -\frac{1}{t_{i+1}-t_i}.
\end{aligned}$$

It is easy to show that in this case $\left(\mathbf{U}'\mathbf{W}_n\mathbf{U}\right)^{-1}$ is a diagonal matrix with entries $t_2-t_1, t_3-t_2, \ldots, t_n-t_{n-1}$. Thus

$$-\frac{1}{2\sigma_w^2}\mathbf{f}'\mathbf{U}\left(\mathbf{U}'\mathbf{W}_n\mathbf{U}\right)^{-1}\mathbf{U}'\mathbf{f} = -\frac{1}{2\sigma_w^2}\sum_{i=2}^{n}(t_i-t_{i-1})\left(\frac{f(t_i)-f(t_{i-1})}{t_i-t_{i-1}}\right)^2.$$

Well known results on spline interpolation (summarized, for example, in Schoenberg, 1964) imply that the solution to (22), $f(t)$, is a piecewise linear function with knots at $t_1,\ldots,t_n$. Thus we obtain

$$-\frac{1}{2\sigma_w^2}\mathbf{f}'\mathbf{U}\left(\mathbf{U}'\mathbf{W}_n\mathbf{U}\right)^{-1}\mathbf{U}'\mathbf{f} = -\frac{1}{2\sigma_w^2}\int_0^b \left[f'(t)\right]^2 \mathrm{d}t.$$

If we set $\lambda_1 = \lambda / \left(2\sigma_w^2\right)$ the maximisation problem is equivalent to minimising (22) with respect to $\mathbf{f}$ for $m = 1$.

The proof for $m = 2$ proceeds along the same lines. Here we set

$$\begin{aligned}
\mathbf{U}' &= \begin{bmatrix} \mathbf{u}_1' \\ \mathbf{u}_2' \\ \vdots \\ \mathbf{u}_{n-2}' \end{bmatrix}, \quad \mathbf{u}_i' = \left(\underbrace{0, \ldots, 0}_{\text{i-1 zeros}}, \alpha_0^i, \alpha_1^i, \alpha_2^i, \underbrace{0, \ldots, 0}_{\text{n-2-i zeros}} \right), \\
\alpha_0^i &= \frac{1}{t_{i+1} - t_i} \\
\alpha_1^i &= -\frac{1}{t_{i+2} - t_{i+1}} - \frac{1}{t_{i+1} - t_i} \\
\alpha_2^i &= \frac{1}{t_{i+2} - t_{i+1}}.
\end{aligned}$$

It can be shown that $\mathbf{W}_n$ has entries

$$[\mathbf{W}_n]_{ij} = \frac{1}{3}\left[\min\left(t_i, t_j\right)\right]^3 + \frac{1}{2}\left|t_i - t_j\right|\left[\max\left(t_i, t_j\right)\right]^2.$$

Bosch et al. (1995) showed that

$$\lambda_m \int_0^b \left[f''\left(t\right)\right]^2 \mathrm{d}t = \lambda_m \mathbf{f}'\mathbf{U}\left(\mathbf{U}'\mathbf{W}_n\mathbf{U}\right)^{-1}\mathbf{U}'\mathbf{f}$$

where $f(t)$ is the solution to problem (22), a cubic spline with knots at $t_1, \ldots, t_n$. Thus finding the mode is equivalent to minimising (22) with respect to $\mathbf{f}$ for $m = 2$ and $\lambda_2 = \lambda / \left(2\sigma_w^2\right)$.

Acknowledgements Early versions of this paper were presented at the Econometrics in Rio conference in July 2006 and at the 4th Oxmetrics Users Conference, CASS Business School, London in September, 2006. We would like to thank David Hendry, James Davidson, Søren Johansen, Siem Jan Koopman and Roger Koenker for helpful comments. We would like to thank the ESRC for financial support under the grant Time-Varying Quantiles, RES-062-23-0129.

References

Bosch, R.J., Y. Ye and G.G. Woodworth, 1995, A convergent algorithm for quantile regression with smoothing splines, Computational Statistics and Data Analysis 19, 613–630.

De Rossi, G. and A.C. Harvey, 2006, Time-varying quantiles, Faculty of Economics, Cambridge, CWPE 0649.

Durbin, J. and S.J. Koopman, 2001, Time series analysis by state space methods (Oxford University Press, Oxford).

Engle, R.F. and S. Manganelli, 2004, CAViaR: Conditional autoregressive value at risk by regression quantiles, Journal of Business and Economic Statistics 22, 367–381.

Granger, C.W.J. and C-Y. Sin, 2000, Modelling the absolute returns of different stock indices: exploring the forecastability of an alternative measure of risk, Journal of Forecasting 19, 277–298.

Harvey, A.C., 1989, Forecasting, structural time series models and the Kalman filter (Cambridge University Press, Cambridge).

Harvey, A.C. and S.J. Koopman, 2000, Signal extraction and the formulation of unobserved components models, Econometrics Journal 3, 84–107.

Kimeldorf, G. and G. Wahba, 1971, Some results on Tchebycheffian spline functions, Journal of Mathematical Analysis and Applications 33, 82–95.

Koenker, R., 2005, Quantile regression (Cambridge University Press, Cambridge).

Kohn, R., C.F. Ansley and C-H. Wong, 1992, Nonparametric spline regression with autoregressive moving average errors, Biometrika 79, 335–346.

Koopman, S.J., N. Shephard and J. Doornik, 1999, Statistical algorithms for models in state space using SsfPack 2.2, Econometrics Journal 2, 113–166.

Newey, W.K. and J.L. Powell, 1987, Asymmetric least squares estimation and testing, Econometrica 55, 819–847.

Schoenberg, I.J., 1964, Spline functions and the problem of graduation, Proceedings of the National Academy of Science 52, 947–950.

Silverman, B.W., 1985, Some aspects of the spline smoothing approach to non-parametric regression curve fitting, Journal of the Royal Statistical Society B 47, 1–52.

Wahba, G., 1978, Improper priors, spline smoothing and the problem of guarding against model errors in regression, Journal of the Royal Statistical Society B 40 (3), 364–372.